External evaluation of a novel prostate cancer risk calculator (ProstateCheck) based on data of the Swiss arm of the ERSPC

Poyet, Cédric ; Wettstein, Marian S ; Lundon, Dara J ; Bhindi, Bimal ; Kulkarni, Girish S ; Saba, Karim ; Sulser, Tullio ; Vickers, A J ; Hermanns, Thomas

Abstract: PURPOSE: To externally validate a novel prostate cancer (PCa) risk-calculator (RC), based on data of the Swiss arm of the European Randomize Study of Screening for Prostate Cancer (ERSPC), and to assess whether the RC (ProstateCheck) is superior to the PCPT RC and SWOP RC in an independent Swiss cohort. MATERIALS AND METHODS: Data of all men who underwent prostate biopsy in an academic tertiary care center between 2004 and 2012 were retrospectively analyzed. The probability of having any PCa or high-grade PCa (Gleason score 7) upon prostate biopsy was calculated using the ProstateCheck-RC. The RC's performance was assessed using calibration and discrimination and additionally compared with the PCPT- and SWOP-RC by decision curve analyses. RESULTS: Of 1615 men, 401 (25%) were diagnosed with any PCa and 196 (12%) with high-grade PCa. Our analyses of the ProstateCheck-RC revealed good calibration in the low risk range (0 to 0.4) and moderate overestimation in the higher risk range (0.4 to 1) for any and high-grade PCa. The AUC for the discrimination of any PCa and high-grade PCa was 0.69 and 0.72, respectively, which was slightly but significantly higher compared to the PCPT-RC (0.66 and 0.69, respectively) and SWOP-RC (0.64 and 0.70, respectively). Decision analysis taking into account the harms of transrectal ultrasound measurements of prostate volume, found little benefit for ProstateCheck-RC, with inferior properties to the PCPT- and SWOP-RC. CONCLUSION: Our independent external evaluation revealed moderate performance of the ProstateCheck-RC. Its clinical benefit is limited and inferior to the PCPT- and SWOP-RC.

DOI: https://doi.org/10.1016/j.juro.2016.05.081

DOI: https://doi.org/10.1016/j.juro.2016.05.081

External evaluation of a novel prostate cancer risk calculator (ProstateCheck) based on data of the Swiss arm of the ERSPC

Cédric Poyet[a†], Marian S. Wettstein[a†], Dara J. Lundon[b], Bimal Bhindi[c], Girish S Kulkarni[c], Karim Saba[a], Tullio Sulser[a], A. J. Vickers[d] and Thomas Hermanns[a*]

[a]Department of Urology, University Hospital Zürich, University of Zürich, Zürich, Switzerland
[b]Department of Urology, Mater Misericordiae University Hospital, Dublin, Ireland
[c]Department of Surgery, Division of Urology, University Health Network, University of Toronto, Toronto, Canada
[d]Department of Epidemiology and Biostatistics, Memorial Sloan Kettering Cancer Center, New York, USA

[*]Corresponding author:
Department of Urology
University Hospital Zürich
Frauenklinikstrasse 10
8091 Zürich, Switzerland
Phone: +41-44-255-1111
Fax: +41-44-255-4555
email: thomas.hermanns@usz.ch

[†]equal contribution

Running head

ProstateCheck risk calculator performance

Keywords

Prostate cancer, biopsy, prostate-specific antigen, nomograms, decision aids

Word count: 2500

Abstract

Purpose: To externally validate a novel prostate cancer (PCa) risk-calculator (RC), based on data of the Swiss arm of the European Randomize Study of Screening for Prostate Cancer (ERSPC), and to assess whether the RC (ProstateCheck) is superior to the PCPT RC and SWOP RC in an independent Swiss cohort.

Materials and Methods: Data of all men who underwent prostate biopsy in an academic tertiary care center between 2004 and 2012 were retrospectively analyzed. The probability of having any PCa or high-grade PCa (Gleason score ≥7) upon prostate biopsy was calculated using the ProstateCheck-RC. The RC's performance was assessed using calibration and discrimination and additionally compared with the PCPT- and SWOP-RC by decision curve analyses.

Results: Of 1615 men, 401 (25%) were diagnosed with any PCa and 196 (12%) with high-grade PCa. Our analyses of the ProstateCheck-RC revealed good calibration in the low risk range (0 to 0.4) and moderate overestimation in the higher risk range (0.4 to 1) for any and high-grade PCa. The AUC for the discrimination of any PCa and high-grade PCa was 0.69 and 0.72, respectively, which was **slightly but significantly** higher compared to the PCPT-RC (0.66 and 0.69, respectively) and SWOP-RC (0.64 and 0.70, respectively). Decision analysis taking into account the harms of transrectal ultrasound measurements of prostate volume, found little benefit for ProstateCheck-RC, with inferior properties to the PCPT- and SWOP-RC.

Conclusion: Our independent external evaluation revealed moderate performance of the ProstateCheck-RC. Its clinical benefit is limited and inferior to the PCPT- and SWOP-RC.

Introduction

Numerous risk calculators (RCs) have been developed to predict prostate cancer (PCa) detection on transrectal biopsy.[1-4] These multivariable risk-prediction tools have been shown to be superior to prediction using prostate-specific antigen (PSA) testing and digital rectal examination (DRE) alone.[2,5] The use of PCa RCs is increasingly recommended to reduce overdiagnosis and eventually overtreatment, which are a result of PSA screening without personalized risk assessment.[6-8]

Original reports of novel RCs usually show excellent performance.[2,9,10] However, in external validations RC performance is often less optimistic.[11-16] There are multiple reasons for differences in RC performance in different settings. Some RCs are based on prospective or retrospective analyses from clinical biopsy series[2,17] while others are based on data from randomized studies.[1,4] Therefore, the underlying patient cohorts often do not reflect real-life scenarios of individualized screening. Differences in biopsy strategies (e.g. core numbers) or tested populations (e.g. ethnic, environmental and geographical) may also account for differences in RC performance.

A regionally developed RC might thus have advantages over RCs that were developed using cohorts of different geographical regions or ethnic compositions. However, whether a regionally developed RC really performs better is unknown.

The aim of the present study was to evaluate the performance of a recently launched RC (ProstateCheck) developed from data of the Swiss arm of the European Randomized Study for Screening of PCa (ERSPC Aarau) in a contemporary independent Swiss cohort.[18,19]

Material and Methods

All men who underwent a transrectal ultrasound (TRUS)-guided prostate biopsy between January 2004 and July 2012 in a Swiss tertiary care academic center were retrospectively identified. Exclusion criteria were age <45 years or >75 years, history of a previous positive biopsy and serum PSA>50 ng/ml. Men with prior negative biopsies were included. The study was approved by the local ethics committee.

The decision to undergo a prostate biopsy was made in men with a PSA value >2.5ng/ml or an abnormal DRE after individual discussion of potential benefits and harms of the procedure. Before 2007, either six- or eight-core biopsies and after 2007 only twelve-core biopsies were done. Clinical (pre-biopsy PSA value, age, family history, DRE results (suspicious versus unsuspicious for PCa) and TRUS-measured prostate volume) and pathological data (diagnosis of any PCa and high-grade PCa (defined as Gleason grade ≥7)) were retrieved from electronic hospital charts.

The ProstateCheck-RC is based on data of approximately 5000 men who participated in the Swiss arm of the ERSPC.[20]

Six different parameters (total PSA, ratio of free to total PSA, age, family history, suspicious DRE and prostate volume) can be entered into the RC.

Depending on the PSA value of a patient, the RC calculates different probabilities: For men with a PSA <3.0ng/ml, the RC does not provide the risks at present but gives predictions of PCa diagnosis in the future (i.e. after 4, 8 and 12 years) and recommendations for the timing of follow-up screening investigations (i.e. 1 year, 2-3 or 6-8 years). For men with a PSA ≥3ng/ml the RC predicts the current risk of having any PCa and high-grade PCa. Recommendations for further urological workup are given according to the calculated risk of any PCa.

For this study, patients with a PSA value ≥3ng/ml were further analyzed, as the RC only predicts current PCa risk when the PSA values is ≥3ng/ml. The probabilities for any PCa and high-grade PCa were calculated and compared with the actual biopsy results of the entire cohort.

Risk calculation for our analysis was performed in June 2015 using the ProstateCheck version 1.2. All parameters except free/total PSA ratio, which was not available in our dataset **but was also not mandatory for PCa risk prediction**, **were manually entered into the smartphone application for every patient.**

Calibration and discrimination of the RC were evaluated as previously described.[16] Calibration of the RC was analyzed graphically using calibration plots. Furthermore, the calibration slope was calculated and calibration-in-the-large was assessed. Discrimination was evaluated using receiver-operation characteristic (ROC) curves and the area under the curve (AUC) with corresponding 95% confidence intervals (CI). **The AUCs of the different RCs were compared using the DeLong test.**

To investigate whether the number of biopsy cores taken **or a prior negative biopsy** had an impact on the performance of the RC, sensitivity analys**e**s for patients with 6-8-core biopsies versus 12-core biopsies **and for biopsy-naïve men versus men who previously had a negative biopsy** were performed.

Finally, the ProstateCheck-RC was compared with two well-known RCs: the PCPT-RC version 2.0 and the SWOP-RC version 3+DRE/4+DRE.[1,9,10] These two RCs have previously been validated and compared in our cohort.[16] For the SWOP-RC the TRUS volumes of our cohort had to be trichotomized (for details see [16]). In the present study only patients with a PSA value ≥ 3ng/ml were analyzed to allow comparison with the ProstateCheck-RC.

Decision curve analyses (DCAs) for the prediction of high-grade PCa were performed to assess the net benefit of the three different RCs according to different threshold probabilities

at which one would consider to perform a biopsy.[21] In contrast to the SWOP-RC, which includes prostate volume estimation by DRE and the PCPT-RC, which does not include prostate volume at all, the ProstateCheck-RC requires prostate volume measured by TRUS. The harm of this unpleasant and costly investigation needs to be taken into account when evaluating the RC. Net harm, which is subtracted from net benefit, is defined by the reciprocal value of the maximum number of TRUS that one would undertake to detect one high-grade PCa, if TRUS was perfectly accurate for diagnosis high-grade PCa. In our base case, we made the liberal assumption that a clinician would be prepared to subject as many as 50 patients to TRUS in order to find one high-grade cancer, but we varied this assumption in sensitivity analyses.

Statistical analyses were performed using the IBM SPSS Statistics version 22 (IBM, Armonk, USA) and R programming language and software environment version 3.1.3 (R Foundation for Statistical Computing, Vienna, A) with the extension packages of Kundu and colleagues for the calibration plots[22] and Vickers and colleagues for the DCAs (http://www.decisioncurveanalysis.org).[21]

Results

Of 2304 men 689 (29.9%) were excluded because they were younger than 45 years or older than 75 years, had a PSA <3ng/ml or >50ng/ml or had a previous positive biopsy. Baseline characteristics and biopsy results of the final cohort (*n*=1615) are summarized in Table 1. Almost all men were of Caucasian descent (99%). Any PCa and high-grade PCa were found in 401 men (25%) and 196 men (12%), respectively.

Analyses for calibration-in-the-large of the ProstateCheck-RC showed a predicted rate of 23% for any PCa compared to an actual detection rate of 25%. The predicted rate for high-grade PCa (11%) was also comparable to the actual detection rate (12%; Table 2).

The calibration plots revealed good calibration in the low risk range (0 to 0.4) and moderate overestimation in the higher risk range (0.4 to 1) for any and high-grade PCa (Figure 1,2).

Analyses of the discriminative ability gave an AUC of 0.69 (95%CI=0.67-0.73) for any PCa and 0.72 (95%CI=0.69-0.77) for high-grade PCa (Table 2). The AUCs of PSA alone were 0.58 (95%CI=0.55-0.61) for any PCa and 0.63 (95% CI=0.59-0.67) for high-grade PCa.

Our sensitivity analysis stratifying by biopsy core number revealed a higher detection rate after 12-core biopsies than after 6-8-core biopsies for any PCa (28.4% vs. 18.9%) and high-grade PCa (17% vs. 6.8%; Supplemental Table1). The RC overestimated the risk for any and high-grade PCa in the 6-8-core group. In the 12-core group the prediction of any PCa was close to the observed rate but the RC underestimated the risk for high-grade PCa. The AUCs were lower in the 12-core group compared to the 6-8-core group for any PCa (0.68 vs. 0.72) and high-grade PCa (0.71 vs. 0.79).

Our second sensitivity analysis revealed a higher detection rate in the biopsy-naïve group for any PCa (30.8% vs. 17.4%) and high-grade PCa (16.9% vs. 6.2%; Supplemental Table 2). The RC overestimated the risk for any and high-grade PCa in

the group of men with a previous negative biopsy. For biopsy-naïve men, the estimates were close to the observed proportion.

For the PCPT- and SWOP-RC the predicted and actual detection rates for any and high-grade PCa were similar (Table 2). The calibration slopes of the ProstateCheck- and SWOP-RC were comparable for any PCa and somewhat higher for the ProstateCheck-RC for high-grade PCa (Table 2). The calibration plots for the PCPT-RC revealed good calibration in the whole prediction range for any and high-grade PCa (Supplemental Figures 1a,b). For the SWOP-RC, calibration plots revealed reasonable calibration with moderate overestimation for any PCa and mild underestimation for high-grade PCa in the higher risk ranges (Supplemental Figures 2a,b).

Compared to the ProstateCheck-RC, the AUCs of the PCPT- and SWOP-RCs were lower for any PCa (0.66 and 0.64, respectively) and high-grade PCa (0.69 and 0.70 respectively, Table 2). **The AUC for high-grade PCa of the ProstateCheck-RC was significantly superior to the PCPT- (P<0.001) and the SWOP-RC (p=0.02).**

The DCAs revealed that the ProstateCheck-RC has limited net benefit and is inferior to the PCPT- and SWOP-RC (Figure 3). Even if one would be prepared to subject 100 men to TRUS in order to find one high-grade cancer, the ProstateCheck-RC was inferior to the other RCs for almost the complete range of threshold probabilities (data not shown).

Discussion

Numerous reasons have been proposed for the limitations of a 'one size fits all' RC including changes in clinical practice over time and differences in patient population between RC development cohorts and tested cohorts.[2,11,23-25] In the present study we independently evaluated the performance of the ProstateCheck-RC, in a contemporary clinical Swiss cohort. The regionally developed ProstateCheck-RC showed fair to moderate performance compared to the two validated international SWOP- and PCPT-RCs. However, decision analysis that incorporated the harm of TRUS-measured prostate volume demonstrated that the ProstateCheck-RC is clearly inferior to the PCPT-RC and SWOP-RC.

The ProstateCheck-RC was released as application for smartphones or tablet computers in December 2014. Its use is subject to payment of a fee for either a limited number of calculations (20 calculations for 3.49US$) or unlimited use during a predefined period (e.g. twelve months for 19.99US$). As of yet, detailed data on its development and performance in the development cohort have not been published. Furthermore, t**he source code of the RC is not available** and external validations are lacking. **The user of the ProstateCheck-RC is informed that the RC was developed based on data of the Swiss arm of the ERSPC, but not that he has to pay for a non-validated tool. Missing detail information restricts scientific exchange and improvement of this RC. This is in contrast to the two other RCs, which are both available online and free-of-charge. Additionally, both RCs have been validated among different cohorts because their source code is available.**

It is unique to this RC that it calculates different probabilities depending on the PSA value of a patient and gives recommendations for users whether and when further workup should be performed. Recommendations for further workup in the PSA group ≥3ng/ml are based on the risk of being diagnosed with any PCa. Although the RC calculates the risk of being diagnosed with high-grade PCa, this risk is not incorporated into the recommendation for further

workup. **We believe that a RC should not focus on the detection of low risk disease but reliably predict the risk for high-grade disease to prevent unnecessary biopsies and overdiagnosis of indolent PCa, which usually should not trigger active treatment but surveillance.**[26,27]

Although the ProstateCheck-RC was developed based on data of a study which was performed in a centre in close vicinity to our centre, we did not identify a clinically meaningful superiority of this regional RC over the two international RCs. Wide variations of PCa incidence have been reported among different countries and regions.[28] These variations may influence RC performance in different cohorts but are expected to be similar in our cohort and the ProstateCheck-RC development cohort. However, there are also important differences between the two cohorts. The RC was developed from data of a population-based mass screening trial. Men in our cohort were biopsied after individualized screening. The underlying biopsy strategy of the RC (6-core biopsy) is different to current recommendations for prostate biopsies (e.g. 10-12-core biopsy), which were followed in most patients in our cohort. Our sensitivity analysis confirmed an inferior performance of the ProstateCheck-RC in the 12-core biopsy group. Furthermore, our cohort, as many clinical cohorts, included a notable amount of previously biopsied men. A prior negative biopsy, which **has an impact on RC performance, as shown in our sensitivity analysis,** is a variable of the SWOP-RC and PCPT-RC but not of the ProstateCheck-RC.[9,10] Strobl and colleagues found that recalibration according to local cohort characteristics can improve the performance of a static RC, that was originally designed as a one-size-fits all model.[23] However, our data indicates that recalibration cannot be evaded by the use of a regional RC.

The ProstateCheck-RC includes the variable TRUS-measured prostate volume. **Interobserver variability of TRUS volume measurement can affect RC performance for an individual patient. Additionally,** TRUS measurements are costly, invasive and

unpleasant. Therefore, to justify its implementation, the investigation must significantly raise the benefit of the RC compared to RCs that do not include this investigation. Our analysis shows that subtraction of potential harm of TRUS measurements significantly reduces the benefit of the RC resulting in superior benefit of the SWOP-RC and PCPT-RC.

Free/total PSA ratio was not available in our dataset, **which we consider the main limitation of this work. The missing variable** might have an impact on the results of our validation. As the full details of the weight of different variables of the RC are not yet published, it remains unknown what the impact of this missing value is. Additionally prostate volumes for the SWOP-RC were not based on DRE but on TRUS measurements, which are known to be more accurate. This might lead to a better performance of the RC in the present investigation than expected if true DRE-based prostate volume was used. However, the performance of the PCPT-RC, which does not include prostate volume and thus had the most robust performance in our investigation, confirms the inferiority of the ProstateCheck-RC.

Our results nicely illustrate that the ProstateCheck-RC does not have relevant benefit over the pre-existing and well-evaluated RCs. It seems that all RCs have limitations in contemporary patient cohorts due to the above-mentioned aspects. Thus, future research should rather focus on the development of RCs, which **implement** modern state of the art biopsy strategies and results from imaging studies such as multiparametric MRI [29] or novel biomarkers. The updated version of the SWOP RC including the Prostate Health Index (PHI) has recently been shown to be superior to the conventional SWOP-RC in an independent validation study.[30]

Conclusion

Our independent, external evaluation revealed that the novel ProstateCheck-RC adds no benefit to PCa risk prediction compared to the PCPT or SWOP-RC. Future research should focus on RCs, which are based on contemporary clinical cohorts including novel tools such as biomarkers or results from imaging studies.

References

1. R. Kranse, M. Roobol and F.H. Schroder, A graphical device to represent the outcomes of a logistic regression analysis, Prostate **68** (2008), p. 1674.

2. R.K. Nam, M.W. Kattan, J.L. Chin *et al.*, Prospective multi-institutional study evaluating the performance of prostate cancer risk calculators, J Clin Oncol **29** (2011), p. 2959.

3. M.J. Roobol, E.W. Steyerberg, R. Kranse *et al.*, A risk-based strategy improves prostate-specific antigen-driven detection of prostate cancer, Eur Urol **57** (2010), p. 79.

4. I.M. Thompson, D.P. Ankerst, C. Chi *et al.*, Assessing prostate cancer risk: results from the Prostate Cancer Prevention Trial, J Natl Cancer Inst **98** (2006), p. 529.

5. F. Schroder and M.W. Kattan, The comparability of models for predicting the risk of a positive prostate biopsy with prostate-specific antigen alone: a systematic review, Eur Urol **54** (2008), p. 274.

6. I.M. Thompson, Jr., R.J. Leach and D.P. Ankerst, Focusing PSA testing on detection of high-risk prostate cancers by incorporating patient preferences into decision making, JAMA **312** (2014), p. 995.

7. M.J. Roobol, Prostate cancer: Rescreening policies and risk calculators, Nat Rev Urol **11** (2014), p. 429.

8. A. Heidenreich, P.A. Abrahamsson, W. Artibani *et al.*, Early detection of prostate cancer: European association of urology recommendation, Eur Urol **64** (2013), p. 347.

9. D.P. Ankerst, J. Hoefler, S. Bock *et al.*, Prostate Cancer Prevention Trial risk calculator 2.0 for the prediction of low- vs high-grade prostate cancer, Urology **83** (2014), p. 1362.

10. M.J. Roobol, H.A. van Vugt, S. Loeb *et al.*, Prediction of prostate cancer risk: the role of prostate volume and digital rectal examination in the ERSPC risk calculators, Eur Urol **61** (2012), p. 577.

11. D.P. Ankerst, A. Boeck, S.J. Freedland *et al.*, Evaluating the PCPT risk calculator in ten international biopsy cohorts: results from the Prostate Biopsy Collaborative Group, World J Urol **30** (2012), p. 181.

12. V. Cavadas, L. Osorio, F. Sabell *et al.*, Prostate cancer prevention trial and European randomized study of screening for prostate cancer risk calculators: a performance comparison in a contemporary screened cohort, Eur Urol **58** (2010), p. 551.

13. D.J. Lundon, B.D. Kelly, R. Foley *et al.*, Prostate cancer risk assessment tools in an unscreened population, World J Urol (2014), p.

14. M. Oliveira, V. Marques, A.P. Carvalho *et al.*, Head-to-head comparison of two online nomograms for prostate biopsy outcome prediction, BJU Int **107** (2011), p. 1780.

15. G. Trottier, M.J. Roobol, N. Lawrentschuk *et al.*, Comparison of risk calculators from the Prostate Cancer Prevention Trial and the European Randomized Study of Screening for Prostate Cancer in a contemporary Canadian cohort, BJU Int **108** (2011), p. E237.

16. C. Poyet, D. Nieboer, B. Bhindi *et al.*, Prostate cancer risk prediction using the novel versions of the ERSPC and PCPT risk calculators: Independent validation and comparison in a contemporary European cohort, BJU Int (2015), p.

17. P.I. Karakiewicz, S. Benayoun, M.W. Kattan *et al.*, Development and validation of a nomogram predicting the outcome of prostate biopsy based on patient age, digital rectal examination and serum prostate specific antigen, J Urol **173** (2005), p. 1930.

18. M. Kwiatkowski, A. Huber, B. Stamm *et al.*, Features and preliminary results of prostate cancer screening in Canton Aargau, Switzerland, BJU Int **92 Suppl 2** (2003), p. 44.

19. M. Randazzo, S. Wyler, A. Huber *et al.*, ProstateCheck©: Entwicklung des Schweizer Risikokalkulators für ein Finetuning im Prostatakarzinom-Screening, 67. Kongress der Deutschen Gesellschaft für Urologie e.V. 2015 (2015), p.

20. F. Recker and M. Kwiatkowski, App berechnet in Kürze persönliche Prostatagesundheit. http://www.prostatakrebs.ch, (2015), p.

21. A.J. Vickers and E.B. Elkin, Decision curve analysis: a novel method for evaluating prediction models, Med Decis Making **26** (2006), p. 565.

22. S. Kundu, Y.S. Aulchenko, C.M. van Duijn *et al.*, PredictABEL: an R package for the assessment of risk prediction models, Eur J Epidemiol **26** (2011), p. 261.

23. A.N. Strobl, A.J. Vickers, B. Van Calster *et al.*, Improving patient prostate cancer risk assessment: Moving from static, globally-applied to dynamic, practice-specific risk calculators, J Biomed Inform **56** (2015), p. 87.

24. D.P. Ankerst, A. Boeck, S.J. Freedland *et al.*, Evaluating the Prostate Cancer Prevention Trial High Grade Prostate Cancer Risk Calculator in 10 international biopsy cohorts: results from the Prostate Biopsy Collaborative Group, World J Urol **32** (2014), p. 185.

25. M.W. Kattan, Factors affecting the accuracy of prediction models limit the comparison of rival prediction models when applied to separate data sets, Eur Urol **59** (2011), p. 566.

26. R.A. Godtman, E. Holmberg, A. Khatami *et al.*, Outcome following active surveillance of men with screen-detected prostate cancer. Results from the

Goteborg randomised population-based prostate cancer screening trial, Eur Urol **63** (2013), p. 101.

27. J.H. Hayes, D.A. Ollendorf, S.D. Pearson *et al.*, Observation versus initial treatment for men with localized, low-risk prostate cancer: a cost-effectiveness analysis, Ann Intern Med **158** (2013), p. 853.

28. M.M. Center, A. Jemal, J. Lortet-Tieulent *et al.*, International variation in prostate cancer incidence and mortality rates, Eur Urol **61** (2012), p. 1079.

29. S.S. Salami, M.A. Vira, B. Turkbey *et al.*, Multiparametric magnetic resonance imaging outperforms the Prostate Cancer Prevention Trial risk calculator in predicting clinically significant prostate cancer, Cancer **120** (2014), p. 2876.

30. R.W. Foley, R.M. Maweni, L. Gorman *et al.*, The ERSPC Risk Calculators Significantly Outperform The PCPT 2.0 In The Prediction Of Prostate Cancer; A Multi-Institutional Study, BJU Int (2016), p.

Figure legends

Figure 1 and 2: Calibration plots for the ProstateCheck-RC predicting any prostate cancer (1) and high-grade prostate cancer (2). The x-axis shows predicted probabilities by the models and y-axis shows observed deciles. Deciles are shown by bullets with 95% Confidence intervals.

Figure 3: Decision curve analysis of predicting high-grade prostate cancer on biopsy using the ProstateCheck-RC (green dashed line), the SWOP RC (black dashed line) and the PCPT RC (red dashed line). For the ProstateCheck-RC the net harm of TRUS measurements of prostate volume was incorporated into the decision curve. We assumed that a clinician would do no more than 50 TRUS to detect one high-grade PCa, even if TRUS were a perfect diagnostic test.